ELSA: Entropy-based Local Indicator of Spatial Association

Babak Naimi\textsuperscript{1,2*}, Nicholas A.S. Hamm\textsuperscript{3}, Thomas A. Groen\textsuperscript{4}, Andrew K. Skidmore\textsuperscript{4,5}, Albertus G. Toxopeus\textsuperscript{4}, Sara Alibakhshi\textsuperscript{5}

\textsuperscript{1} Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands;
\textsuperscript{2} Department for Migration and Immuno-ecology, Max Planck Institute for Ornithology, Am Obstberg 1, 78315, Radolfzell, Germany;
\textsuperscript{3} School of Geographical Sciences and Geospatial Research Group, Faculty of Science and Engineering, University of Nottingham, Ningbo, China;
\textsuperscript{4} Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands;
\textsuperscript{5} Department of Environmental Science, Macquarie University, NSW 2109, Australia

\textsuperscript{6} Department of Built Environment, School of Engineering, Aalto University, P.O. Box 14100, 00076 Espoo, Finland

\textsuperscript{*} Correspondence: Babak Naimi, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands; Email: naimi.b@gmail.com

Highlights

- ELSA is a new local statistic can be used for both categorical and continuous spatial data.
- ELSA quantifies the degree of local spatial association of a variable at each location.
- Entrogram quantifies global spatial structure and represent as a variogram-like graph.
- ELSA is a non-parametric reliable and robust statistic.
- The R package, elsa, provides the tools for calculation of ELSA and entrogram.
Abstract

Research on spatial data analysis has developed a number of local indicators of spatial association (LISA), which allow exploration of local patterns in spatial data. These include local Moran’s $I$ and local Geary’s $c$, as well as $G_i$ and $G_i^*$ that can be used for continuous or interval variables only.

Despite numerous situations where qualitative (nominal/categorical) variables are encountered, few attempts have been devoted to the development of methods to explore the local spatial pattern in categorical data. To our knowledge, there is no indicator of local spatial association that can be used for both continuous and categorical data at the same time.

In this paper, we propose a new local indicator of spatial association, called the entropy-based local indicator of spatial association (ELSA), can be used for both categorical and continuous spatial data. ELSA quantifies the degree of spatial association of a variable at each location relative to the same variable at the neighbouring locations. This indicator simultaneously incorporates both spatial and attribute aspects of spatial association into account. The values of ELSA vary between 0 and 1, which denote highest and lowest spatial association, respectively. We compare ELSA to existing statistics such as Local Moran’s $I$ and test the power and size of the new statistic. We also introduce the "entrogram", a novel approach for exploring the global spatial structure within the entire area (like a variogram). This study showed that the ELSA is consistent and robust, and is therefore suitable for applications in a wide range of disciplines. The ELSA algorithm is made available as an R-package (elsa).
1. Introduction

Spatial analysis is concerned with exploration and identification of associations over geographical space. Such associations quantify the degree to which a value of a variable measured at one location is dependent on the values of the same variable measured at a specific geographic distance from that location (Cliff and Ord 1981; Goodchild 1986). If such dependency exists in a dataset, the variable is said to exhibit spatial autocorrelation (Sokal and Oden 1978). Several statistics have been developed to quantify spatial autocorrelation both globally and locally. Global measures provide a statistic for the entire field under the assumption of spatial stationarity, i.e. the mean and covariance do not vary over space. This assumption is often unrealistic. Recent advances have addressed non-stationarity in the mean through spatially varying coefficient modelling (Finley 2011; Gelfand et al. 2003; Hamm et al. 2015a). Heteroskedasticity in the variance can be addressed through a weighting function (Hamm et al. 2012; Lark 2009) and further efforts have been directed at modelling non-stationary covariance functions (Haskard and Lark 2009; Paciorek and Schervish 2006). Models that address non-stationary are often difficult to implement and there is a lack of standard software tools. Further, these models all provide a global measure of autocorrelation. This may be of limited relevance when local areas are of interest.

Research studies on spatial data analysis developed a number of local spatial statistics (Anselin 1995; Boots 2003; Getis and Ord 1992; Ord and Getis 1995). In contrast to global measures, these statistics allow exploration of local patterns in spatial association (Lloyd 2007). They do not rely on the assumption of global stationarity. Anselin (1995) introduced a set of statistics, called local indicators of spatial association (LISA), including local Moran’s $I$ and local Geary’s $c$, that decompose a single global measure into the contribution of each individual location to explore the locations that are major contributors to the global autocorrelation. These statistics can be used to test if local spatial clustering of similar values around the observation is significantly different from the global mean. Getis and Ord (1992) and Ord and Getis (1995) also defined two local statistics, $G_i$ and $G_i^*$,
which are somewhat different from Anselin’s LISAs, indicating local clustering of high and low values.

These allow detection of pockets of spatial association that may not be evident when using global
statistics (Getis and Ord 1992; Ord and Getis 1995). These methods, however, can be used for
continuous or interval variables only. Despite numerous situations where qualitative
(nominal/categorical) variables are encountered, only a few attempts have been devoted to develop
methods to explore the spatial pattern for categorical data. Examples include early works using joint
count statistics (Dacey 1968; Dale 2000; Iyer 1949; Moran 1948), extension of the Moran Coefficient
to a wide variety of probability models and linking continuous and discrete variables (Griffith 2010),
and more recent works including local indicators for categorical data (LICD) by Boots (2003) and
symbolic entropy by Ruiz et al. (2010). The purpose of LICD is to identify the nature and spatial
extent of local neighbourhoods that are distinctive or unusual compared to a priori expectation in
binary categorical data (Boots 2003, 2006). To our knowledge, there is no statistic of local spatial
association that can be used for both continuous and categorical data.

In this paper, we propose a new local indicator of spatial association, called the entropy-based local
indicator of spatial association (ELSA), for exploratory analysis and testing the significance of local
spatial association of both categorical and continuous spatial data. In doing so, we address the gaps
in the research literature highlighted above. Entropy has its root in thermodynamic and information
theory. Entropy based approaches have been applied in many disciplines, as a measure of
complexity in physics (Shannon and Weaver 1949), as a measure of diversity or structural complexity
in ecology (Anand and Orloci 1996; Ricotta and Anand 2006), and as a measure of information
content or uncertainty in information theory (Yeung 2008). This concept has also been used by
geographers, economists and social scientists to describe spatial phenomena (Batty 1974, 1976;
Heikkila and Hu 2006). Recently, some attempts have been conducted to use the concept of entropy
as a measure of spatial contiguity for qualitative data (Ruiz et al. 2010), or to detect spatially varying
multivariate relationships (Guo 2010). Matilla-García et al. (2011) highlighted the potential role that
entropy-based measures might play in detecting spatial structure. They used spatial symbolic
entropy for detecting the order of contiguity (spatial lag) of a spatial dependent process (Matilla-
García and Marín 2011).

Although entropy-based approaches have been used widely for categorical data (e.g. soil type,
classified data), there is a challenge in using entropy for exploratory analysis of continuous data. For
these data, the probability density function is often unknown, which is necessary to calculate
entropy (Guo 2010). A common solution to this problem in most practical applications is to construct
a contingency table by binning (discretizing) a continuous variable into a finite number of classes
(Guo 2003; Journel and Deutsch 1993), which then are used as categorical data in the entropy
measure. This, however, raises two new challenges. Firstly, binning continuous variables causes
information loss in data. Secondly, dissimilarity between binned data is not the same as it is in
categorical data. For example, when using a categorical land use map in spatial analysis, no
difference in the level of dissimilarity between pairs of classes is assumed. If a continuous variable
binned into, for example, five categories (C1, ..., C5), then C1 is more similar to C2 than to C5. In this
paper, we illustrate how ELSA addresses these challenges.

The main objectives of our study are: (1) to introduce a new statistic for measuring and testing local
spatial association (ELSA) that can be used for both continuous and categorical spatial data; (2) to
explore the application of ELSA for calculating local spatial association and comparing this with other
indicators; (3) to demonstrate the usability of ELSA to detect global patterns as well, and calculate a
variogram-like global spatial structure, named ‘entrogram’. In addition, we developed a new
software package, ‘elsa’, in the R environment of statistical computing (R Development Core Team
2017) to freely provide the tools required for calculation of the ELSA statistic and the entrogram.
2. Entropy

In information theory, the concept of complexity is closely related to predictability. It is the amount of information required to achieve an optimal prediction (Boschetti 2008). The entropy measure is also known as the information content. The Shannon entropy has been defined as an average amount of information to eliminate uncertainty, given by a finite number of events:

\[ H = - \sum_{k=1}^{m} p_k \log_2 p_k \]  

(5)

where \( H \) measures the entropy of a system with a finite number of \( m \) possible events, and \( p_k \) represents the probability of event \( k \). \( H \) is at a maximum when all events occur in equal abundance and can be quantified by \( \log_2 m \). This measure can be standardized by dividing by \( \log_2 m \), providing a measure of relative entropy ranging between 0 and 1. The function is dimensionless and depends only upon the number of events, not upon any other invariant property of the system to which it is applied (Batty 1976).

3. ELSA statistic

The Entropy based Local indicator of Spatial Association, ELSA, extends the above described entropy measure using a term that summarizes the attribute distance between a location and its neighbourhood locations over a given geographical distance.

Assume \( x = (x_1, x_2, \ldots, x_n)' \) are \( n \) observations related to a spatial process at locations \( u = (u_1, u_2, \ldots, u_n)' \). Further, denote by \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m) \) the set of possible categories that \( x_i \) can take. For a categorical variable, it is usually assumed that pairs of categories are equally dissimilar. There are, however, situations where the level of dissimilarity varies for different pairs of categories. For example, 'dense forest' and 'sparse forest' in a land use map are more similar than a pair of either of
these two classes and 'lake' (for more details, see the Section "ELSA for categorical data"). Likewise,
when the values of a continuous variable are binned (discretized) into categories, the level of
dissimilarity varies between categories. By ranking the binned values, the difference between rank
numbers can be interpreted as the level of dissimilarity. These levels of dissimilarity between
categories are taken into account in the calculation of ELSA.

ELSA (E statistic) at site \( i \) is defined as:

\[
E_i = E_{ai} \times E_{ci}
\]

\[
E_{ai} = \frac{\sum_{j} \omega_{ij} d_{ij}}{\max \{d\} \sum_{j} \omega_{ij}}, \quad j \neq i
\]

\[
E_{ci} = -\frac{\sum_{k=1}^{m} p_k \log_2 (p_k)}{\log_2 m_i}, \quad j \neq i
\]

\[
m_i = \begin{cases} 
  m & \text{if } \sum_{j} \omega_{ij} > m \\
  \sum_{j} \omega_{ij}, & \text{otherwise}
\end{cases}
\]

\[
d_{ij} = |c_i - c_j|
\]

where \( \omega_{ij} \) is a binary weight which specifies whether the site \( j \) is within a specified distance (defines
the neighbourhood size) from site \( i \). \( d_{ij} \) describes the dissimilarity between \( x_i \) and \( x_j \), which is
calculated as the absolute difference of the ranks assigned to the categories at sites \( i \) and \( j \) (i.e., \( c_i \)
and \( c_j \)), and \( \max \{d\} \) is the maximum possible dissimilarity between any pair of observations in the
entire dataset. This is discussed for continuous and categorical variables in the upcoming sections.

There are \( m \) categories in the entire dataset, \( p_k \) is the probability of \( k^{th} \) category from the \( m_o \)
categories within the local distance from site $i$, and $m_i$ is the maximum possible number of
categories within the local distance from site $i$. This means that if the number of observations within
the local distance from site $i$, including site $i$, is greater than the number of categories in the entire
dataset ($\sum_j \omega_{ij} > m$), then $m_i$ is equal to the number of categories, otherwise it is equal to the
number of observations ($\sum_j \omega_{ij}$) within the local distance from site $i$.

The first term in the equation 6 ($E_{ai}$) summarizes the attribute distance (dissimilarity) between site $i$
and the neighbouring sites. This coefficient is bounded between 0 and 1. Low values indicate high
similarity of site $i$ to neighbouring sites, and high values indicate low similarity with neighbouring
sites.

The second term of the $E_i$ statistic (i.e., $E_{ci}$) is the Shannon entropy (equation 5), normalized by
$\log_2 (m_i)$. This term ranges between 0 and 1. By normalizing, values are invariant to the number of
categories present in a dataset. In other words, datasets with different numbers of categories are
comparable. For normalizing, $m_i$ is defined in relation to the global number of categories in the
entire dataset. This term quantifies composition or diversity of the categories within the local
distance from site $i$, but it is not sensitive to the level of dissimilarities between pairs of
observations.

Fig. 1 shows the behaviour of the two terms in ELSA as well as ELSA itself at site $i$ in the centre of a 3
x 3 window under five scenarios. When a location is surrounded by similar locations (example a),
both the dissimilarity and composition statistics (i.e., $E_{ai}$ and $E_{ci}$, respectively) would be calculated
as 0. The composition in the binary maps in examples b and c, is the same, while the dissimilarity
between site $i$ and its neighbour locations in the two maps is different. In example b, the site is
surrounded by 8 dissimilar locations to the site, while in the example c, the site is surrounded by 7
similar and only one dissimilar locations. Therefore, $E_{ai}$ is at its maximum value in example b, while
it is much lower in example c. In the two maps (examples d and e) with three categories, the
dissimilarity is the same for both as the site \( i \) is surrounded by 8 dissimilar locations to the site. The composition, however, is different in these two maps.

These five examples show the complementarity of the two terms in the ELSA statistic in different situations. The role of these terms would even be more important if the categories have different levels of dissimilarity (see the upcoming section). It also suggests that either of these terms would be informative on their own for the purpose they are designed for. For instance, if one needs to identify locations with extreme or outlier values, compared to their neighbour locations, the \( E_{ai} \) term in ELSA would be the statistic that fulfils the need, while if only the composition (i.e., diversity) of the values within a local neighbourhood is the purpose of the study, the \( E_{ci} \) term in ELSA can be used.

3.1. ELSA for categorical data

Categorical variables are typically conceptualized as having no inherent ordering (Ahlqvist and Shortridge 2010), which means that all pairs of categories are equally dissimilar. When using this simplification of class differences and denoting the dissimilarity as \( d \) (as in equation 6): max \((d) = 1\),

\[
d_{ij} = 1 \text{ if } x_i \neq x_j \text{ and } d_{ij} = 0 \text{ if } x_i = x_j.
\]

The assumption that all categories are equally dissimilar is often an oversimplification. Several studies have been conducted to estimate a measure of dissimilarity between categories (Romme 1982; Uuemaa et al. 2008). Categories may also be classified into a hierarchical structure. For example, the United Nations Food and Agricultural Organization’s (FAO) land cover classification system arranges classes hierarchically (Di Gregorio and Jansen 2009). Such hierarchies may be used to describe the level of dissimilarity. Consider a categorical map with four categories: ‘mixed forest’ \((\alpha_1)\), ‘coniferous forest’ \((\alpha_2)\), ‘olive groves’ \((\alpha_3)\) and ‘vineyards’ \((\alpha_4)\). These can be grouped under
primary categories such that 'mixed forest' and 'coniferous forest' belong to the primary category 'forests' (denoted $\beta_1$) and 'olive groves' and 'vineyards' belong to 'agricultural areas' (denoted $\beta_2$).

To calculate $E_i$ for this categorical map, $c_i$ (the rank number for site $i$) is set to 1 (always), and $c_j$ is set to 1 (if sites $i$ and $j$ are the same), or 2 (if the categories in sites $i$ and $j$ are different, but belong to the same primary category), or 3 (if the categories in sites $i$ and $j$ are different and also belong to different primary categories). Consequently, the level of dissimilarity, $d_{ij}$, between two different subcategories of the same primary category is set as $d_{ij} = 1$ (e.g., $d\left(\alpha_1, \alpha_2 \right) = 1$), and between two different subcategories from different primary categories is set as $d_{ij} = 2$ (e.g., $d\left(\alpha_1, \alpha_4 \right) = 2$).

To develop this algorithm, we specify the dissimilarity for categorical data as

1) If $\alpha_i = \alpha_j$ then $c_i = 1$ & $c_j = 1$ therefore $d_{ij} = 0$

2) If $\left(\alpha_i \neq \alpha_j \right)$ & $\left(\beta_i = \beta_j \right)$ then $c_i = 1$ & $c_j = 2$ therefore $d_{ij} = 1$

3) If $\left(\alpha_i \neq \alpha_j \right)$ & $\left(\beta_i \neq \beta_j \right)$ then $c_i = 1$ & $c_j = 3$ therefore $d_{ij} = 2$

The level of dissimilarity between pairs of categories can be illustrated in a matrix (Table 1). This makes ELSA flexible enough to handle situations where the level of dissimilarity can be specified for pairs of categories or when categories are hierarchically ordered.

The above situation can be extended to the case where there are more levels in the hierarchy. The maximum level of dissimilarity is then equal to the degree of hierarchy. In Fig. 2, the level of dissimilarity is presented schematically for a hierarchical system that contains 3 levels.
3.2. ELSA for continuous data

A key step to calculate ELSA for continuous data is that the variable should be first categorized (binned or discretized) into a number of categories; a procedure that may cause information loss. Inspired by Morrison (1972), we propose an estimation of the optimum number of categories that minimizes the information loss. This optimum is the minimum number of categories that is able to reproduce the spatial data statistically (i.e., that minimizes the loss of information through categorization). To find the optimum number of categories, our procedure uses Spearman’s rank correlation coefficient, $\rho$, as a measure of information between the continuous variable and the categorized variable. If the amount of information is not affected through categorizing, the observed correlation should be equal to one. Any loss of information would result in the observed correlation to be less than one. Therefore, the magnitude of the difference $1 - \rho$ provides a measure of information loss (Quester and Dion 1997). The procedure of selecting the optimum number involves the following steps:

1) The categorization procedure starts with a minimum number of categories $m = 2$

2) The procedure assigns a rank number (between 1 and $m$, where $m$ is the total number of categories) to each category.

3) The $\rho$ coefficient between the continuous values and the assigned ranks is calculated for each iteration.

4) The steps 1 to 3 are repeated, every time by considering one more category (i.e., increasing $m$), until a convergence threshold (e.g., 0.005) is reached. The convergence is defined as the difference between the $\rho$ coefficients of the current and previous iterations.

5) The one-standard-error rule (James et al. 2013) is applied to select the optimum number of categories. First, the standard error of the $\rho$ coefficients is calculated, and then the optimum number of categories would be the lowest number for which the $\rho$ coefficient is within one standard error of the highest $\rho$ coefficient.
It is assumed that the information loss due to the categorization is not substantial when the optimum number is used. Fig. 3, illustrates an example of using this procedure for a continuous variable.

To calculate ELSA for the continuous map using equation 6, the original continuous data \( x \) are then mapped into the ranked categories: \( \alpha = (\alpha_1, \alpha_2, ..., \alpha_m) \) with ranks \( c_1, c_2, ..., c_m \), where \( c_1 = 1 \) and \( c_m = m \). The maximum level of dissimilarity in the entire map is therefore: \( \max (d) = c_m - 1 \) and the dissimilarity between the categories at two locations \( u_i \) and \( u_j \) is \( d_{ij} = |c_i - c_j| \).

4. Inference for ELSA statistic

In this section we propose a non-parametric bootstrap randomization approach to test the local spatial association against a null distribution. The approach is based on repeated resampling from a distribution, \( F_0 \), which satisfies the relevant null hypothesis (Davison and Hinkley 1997). Suppose \( \alpha = (\alpha_1, \alpha_2, ..., \alpha_m) \) are the possible events, outcomes of a spatial process, that \( n \) observations \( x = (x_1, x_2, ..., x_n) \) can take at locations \( u = (u_1, u_2, ..., u_n) \). The null surface can be constructed by rearranging or shuffling the locations (Anselin 1995). Once the null surface \( F_0 \) is constructed, a Monte Carlo simulation with \( R \) runs is used to draw a sample with size \( n \) from the null distribution through a bootstrap resampling procedure (sampling with replacement) for each run. The observed ELSA statistic at site \( i \) \( (E_i) \) can be compared to \( R \) independent values of the statistic obtained from the corresponding samples independently simulated under the null hypothesis (i.e., no spatial autocorrelation). If these simulated values at site \( i \) are denoted by \( E^*_1, ..., E^*_R \), then the probability of accepting the null hypothesis (P-value) at site \( i \) \( (P_i) \) can be approximated by (Davison and Hinkley 1997):
where \( \# \{ E_i \geq E^*_i \} \) indicates number of times the observed ELSA at site \( i \) is greater than or equal to the ELSA values calculated for the bootstrap samples drawn from the null distribution.

4.1. Size and power of the test

We analysed the size and power of our non-parametric test based on the ELSA statistic to investigate whether the test correctly rejects the null hypothesis only when it should be rejected (Bivand 2009). To do so, we conducted a comprehensive set of data simulations to generate both continuous and categorical variables with various levels of no to high spatial autocorrelation. We used an unconditional simulation to construct regular grids of 50 \( \times \) 50 grid cells for each variable. Unconditional simulation is a geostatistical technique that generates a realization of a spatially correlated variable, where the spatial correlation is defined by a variogram (Dungan 1999). The circulant-embedding algorithm (Dietrich and Newsam 1993) implemented in the RandomFields package v. 3.0.44 (Schlather 2009) in the R programming environment, v.3.0.1, was used to conduct the unconditional simulation. The effective variogram range (the maximum geographic separation where two points are expected to be autocorrelated) is determined by the scale parameter, \( \phi \). The variability is determined by the sill, \( C + C_0 \), where \( C \) is the partial sill and \( C_0 \) is the nugget and \( \gamma = C/(C + C_0) \) is the proportion of the variability that has spatial structure. An exponential variogram model with an effective autocorrelation range of 3\( \phi = 50 \) cells and an arbitrary sill of 10 was used for all datasets. We controlled the degree of spatial autocorrelation by changing \( \gamma \) in the variogram models. Five levels of \( \gamma = 0, 0.25, 0.5, 0.75, \) and \( 1 \) were used to define the variogram models (Fig. 4), giving a transition from no to high spatial autocorrelation, respectively. To generate the categorical variable, we followed a second step where the continuous spatially autocorrelated variable \( Y \) was
used to define a discrete spatial process as follows Ruiz et al. (2010). Let \( b_{ik} \) be the \( i^{th} \) breaking point in discretizing \( Y \) into \( k \) categories, and defined as:

\[
p(Y \leq b_{ik}) = \frac{i}{k} \quad \text{with} \quad i < k
\]  

(8)

Let \( A = \{\alpha_1, \alpha_2, ..., \alpha_k\} \) and define the discrete spatial process as:

\[
X_u = \begin{cases} 
\alpha_1 & \text{if} \ Y_u \leq b_{1k} \\
\alpha_i & \text{if} \ b_{i-1k} \leq Y_u \leq b_{ik} \\
\alpha_k & \text{if} \ Y_u \geq b_{k-1k} 
\end{cases}
\]  

(9)

where \( X_u \) is the value of discretized \( Y \) at location \( u \). For each level of spatial autocorrelation, we generated three sets of categorical variables with \( k = 2, 3, \) and 4 categories, and a set of continuous variables (Fig. 5).

For each set of data, 999 replicates were simulated and a test with 999 (=R) runs was applied to each replicate at a level of significance \( \alpha = 0.05 \). We applied the test at the center of the image with different local neighbourhood sizes \( Ne = 1.5, 3, 5, 10, \) and 15 cells. We then recorded the number of times that the null hypothesis is rejected (i.e., when \( P \leq 0.05 \)) to quantify the rejection rate. We would expect that the statistic fails to reject the null hypothesis most of the times when the level of spatial autocorrelation is zero (size of the statistic). At the same time, we would expect that it rejects the null hypothesis more frequently (power of the statistic) as the level of autocorrelation goes up (Ruiz et al. 2010).
The results of the numerical analyses are illustrated in Table 2 for different settings. For all cases, the power and size of the test showed that the statistic performed reasonably well. The power of the test goes up when the level of spatial autocorrelation increases. The power was slightly higher for the continuous variables compared to the categorical variables. The size was also slightly higher for the continuous variables with smaller neighborhood sizes, indicating a slightly greater risk for false positive compared to the categorical variables. The results showed that increasing the number of categories in the categorical variables has less effect on the power and size of the statistic. Not surprisingly, there is a loss in the power of the statistic when the neighborhood size is increased because as the distance is increased, the level of spatial association in decreased. The reason is that as the distance is increased, the level of spatial association is decreased (i.e., increasing the variance in the variogram model).

5. Application of ELSA to assess local spatial association

In this section, we illustrate the use of ELSA by means of several examples using real and synthetic datasets, covering both continuous and categorical variables.

5.1. Experiment with categorical data with the same level of dissimilarity between classes

A land cover map in a raster layer including 2769 grid cells of 1 x 1 km from southern Spain was used for this experiment (Fig. 6-a). The map consists of 6 land cover classes. It is assumed that the level of dissimilarity between pairs of classes is equal. In this experiment, the ELSA was calculated at each cell within a local distance of 5 km (Fig. 6-b). As it was expected, the ELSA statistic becomes 0 for homogenous areas with only one class (mostly at the western part of the area), and when the area
becomes more heterogeneous, the ELSA value becomes higher as it is obvious at the center and eastern parts of the area.

[Fig. 6]

The three specified locations in Fig. 6 (i.e., A, B, and C) and the ELSA values at these locations show how this statistic changes when the landscape changes. By looking at these locations on the land cover map, it can be recognized that the level of heterogeneity changes from low to high from location A to location C.

5.2. Experiment with categorical data with non-equal level of dissimilarity between classes

There are numerous situations where the level of dissimilarity between pairs of classes in a categorical variable is not equal. To show how to deal with these situations, we illustrate two experiments using both synthetic and real data.

-Synthetic data: We generated a synthetic categorical raster consisting of 1024 grid cells (32 rows × 32 columns) and four classes (i.e., A1, A2, B1, and B2). The first and second two classes belong to the main categories of A and B, respectively. So, A1 is more similar to A2 than to B1 or B2. The level of dissimilarity between different pairs from the same main category (i.e., A1-A2 or B1-B2) is set to 1, and from different main categories (e.g., A1-B1, A1-B2, etc.) is set to 2 (Fig. 7).
The region is divided into four zones with different combinations of categories. This shows how ELSA changes under these controlled situations. Zone 1 includes only one class (i.e., A1), Zone 2 includes a random distribution of A1 and A2 (i.e., from the same primary category), Zone 3 includes a random distribution of A1 and B1 (i.e., from two different primary categories), and Zone 4 includes random distribution of all four categories.

The maximum level of dissimilarity ($\max \{d\}$) for this experiment is 2. We calculated ELSA at each grid cell using a $3 \times 3$ window as the local neighbourhood (including diagonal neighbours, i.e., queen's case). We also calculated the mean ELSA for each zone to provide a base for comparison (Fig. 8). The results show that the mean ELSA over the two extreme zones are the minimum (Zone 1) and maximum (Zone 4), respectively. The other two zones both follow the same structure (a random distribution of the two classes, but with a different attribute distances). Since Zone 2 consists of two more similar classes than Zone 3, it is expected that the ELSA statistic for the Zone 2 should be lower than Zone 3. This is backed up by the empirical results (Fig. 8-b and Fig. 8-c).

**[Fig. 8]**

- **Real data:** We used the CORINE [Coordination of Information on the Environment of the European Environmental Agency (EEA 2007)] 2006 land cover map from central Spain including 392336 grid cells of 250 x 250 m. The land cover classes in the map were described using a hierarchical scheme with three levels. The first level indicates the primary land cover class (e.g., agricultural area), which are subdivided to more specific types of land cover classes at the second and third levels (e.g., permanent crops and vineyards at the second and third levels, respectively). A three-digit code is used for each land cover, specifying the class at the three levels from left to right (e.g., 221 and 223.
are ‘vineyards’ and ‘olive groves’ respectively, i.e., class 1 and 3 specified at the third level, respectively, but both belong to the same classes of ‘agricultural areas’ [class 2] at level 1, and ‘Permanent crops’ [class 2] at level 2). A list of the classes at the three levels for the codes is provided in Table A1 in Appendix A.

The categories were ordered hierarchically at the three levels based on the three-digit codes. The attribute distance (level of dissimilarity) between each pair of categories was calculated based on their position on the hierarchical scheme. The maximum level of dissimilarity is 3 (e.g., between class 132 and 211). The dissimilarity between pairs of classes is illustrated in Fig. A1 (the table) & A2 (the hierarchical view) in Appendix A.

We calculated ELSA at each grid cell within a local distance of 5 km (Fig. 10). A visual interpretation of the three specified locations on the land cover map (Fig. 9) show that the local association among the three locations is expected to be minimum and maximum at the locations B and C, respectively. The values of the ELSA map at these locations are consistent with the visual interpretation.

5.3. Experiment with continuous data

We simulated two synthetic continuous rasters with two levels of no and positive spatial autocorrelation. We used unconditional simulation (see Section 4.1) to construct regular grids of 30 × 30 cells for each variable (Fig. 11 & 12). For each raster, we calculated a Moran scatterplot (Anselin 1993), along with global Moran’s I (Moran 1948) to quantify the global spatial
autocorrelation and summarize the overall pattern of spatial structure. We calculated ELSA at each cell given a queen contiguity neighbourhood. For comparison, we also quantified other commonly used local indicators of spatial association; $G^*_c$, local Moran's $I$ and local Geary's $c$. These statistics were compared pairwised to explore to what extent these measures are related. A Spearman statistic was used to test and estimate a rank-based measure of association test between each pair of statistics.

The results from the comparison between different local statistics (Fig. 11 & 12) indicate that when the global spatial autocorrelation is positive, ELSA is only related significantly to local Geary's $c$ ($\rho = 0.675$) while the other local indicators are not related either to ELSA or to each other according to the pairwise comparison. When there is no global spatial autocorrelation, ELSA is related significantly to both local Geary's $c$ and local Moran's $I$. The former is a positive ($\rho = 0.807$), and the later is a negative relationship ($\rho = -0.563$). Local Geary's $c$ is also negatively related to local Moran's $I$ ($\rho = -0.509$). The results showed that the Local $G^*$ statistic is not related to any of the other statistics including ELSA. It has been argued that local statistics are sensitive to the existence of global spatial autocorrelation in the dataset (Ord and Getis 2001), that might be the reason to have varying behaviours between the different cases with the positive and no global spatial autocorrelation. For example, we expect a negative relationship between local Moran's $I$ and local Geary's $c$, that is shown in the results only when there is no global spatial autocorrelation. The relationship between ELSA and local Geary's $c$ is also stronger when there is no autocorrelation.
6. Application of ELSA to assess global spatial structure

In this section we explore if global spatial structure can be assessed by employing ELSA into a procedure of generating a sample variogram-like diagram, called 'entrogram'. The idea is that by assessing ELSA with increasing local distances (lags) and putting the averaged values against these distances in a diagram, we can explore spatial structure. The sample variogram is a well-known approach for exploring spatial structure in continuous variables or binary categorical variables (i.e. indicator variogram; Journel 1983). The semantic sample variogram has been developed for multinomial categorical variables (Ahlqvist and Shortridge 2006). If ELSA can be used for this purpose, then the spatial structure for both continuous and categorical variables can be explored with the same technique. The entrogram for lag distance $h$ is calculated as the mean of the ELSA statistics at different sites within the distance equal to the lag size. The entrogram is calculated as follows:

$$ E(h) = \frac{\sum_{i=1}^{n_h} E_i(h)}{n_h} $$  (7)
where $E(h)$ is the value of the entrogram for distance class $h$, $E_i(h)$ is the ELSA statistic at site $i$ within local distance $h$, $n_h$ is the total number of sites within the distance $h$ for which the ELSA statistic is calculated.

In the following experiments we explored the behaviour of the entrogram for both continuous and categorical data.

### 6.1. Entrogram for categorical data

We generated five synthetic categorical maps, on a 20 x 20 raster grid, to explore the capability of the entrogram for calculating the spatial structure of categorical maps. The first three maps were binary, one with randomly distributed classes, and the other two with spatially structured classes. These binary maps provide the opportunity to compare the entrogram with the existing method of exploring the spatial structure for a binary categorical variable (i.e. an indicator variogram). The two spatially structured binary maps were constructed using an unconditional simulation with spherical variogram models varying in their autocorrelation range. We used the range of 3 and 8 grid cells to construct the maps with relatively low and high spatial autocorrelation, respectively (Fig. 13).

We quantified both the entrogram, using our developed R package (i.e. elsa), and the indicator variogram using the gstat package v. 1.1-3 (Pebesma 2004) in the R development environment (R Development Core Team 2017). For both, we used a lag size equal to one grid cell and the cutoff values (number of lags) equal to 12 grid cells. The graphs are then visually interpreted and compared (Fig. 12). The results showed that the entrograms indicate the same patterns as the indicator variograms, and therefore can be used to interpret global spatial structure in binary categorical maps.
The last two categorical maps were generated with four classes. We assumed that these four classes were equally dissimilar. The classes in the first map were spatially structured, giving a maximum degree of spatial clustering, while in the second map they were distributed randomly. This dataset allowed us to illustrate the capability of the entrogram for exploring the spatial structure of multinomial data. We quantified the entrogram with a lag distance equal to one grid cell and the cutoff value (number of lags) equal to 12 grid cells. The visual interpretation of the graphs (Fig. 14) for the spatially clustered map shows that the mean ELSA value is low within the lower distances and increases when the lag distance increases as would be expected. For the randomly distributed classes, on the other hand, this value remains at the maximum level, showing there is no spatial structure.

6.2. Entrogram for continuous data

We generated five continuous raster maps with different ranges of spatial autocorrelation. We used unconditional simulation to construct regular grids of 200 × 200 cells for each variable. Two variogram models, a Spherical and a Gaussian model, were used with varying values for the sill \( C + C_0 \), nugget \( C_0 \) and scale \( \phi \) parameters for different maps. We assigned \( \phi \) different values including \( \phi = 10 \) and \( \phi = 30 \) grid cells giving relative low and high values for the spatial autocorrelation. A total sill, \( C + C_0 = 10 \), was used for all the datasets, with \( C_0 = 0 \) for two maps and \( C_0 = 3 \) and \( C_0 = 5 \) for the other two maps. Additionally, a white-noise surface (\( \phi = 0 \)) was simulated, giving a total of five maps with different levels of spatial autocorrelation. We then quantified and
visualized the empirical variogram and entrogram for each surface (Fig. 15). The visual comparisons of the graphs showed that the entrogram describes a global spatial structure similar to the empirical variogram.

7. Discussion

ELSA allows for exploring and testing the local spatial association for both categorical and continuous variables. This provides the opportunity of using one statistic for a study where both types of variables are used, e.g., in species distribution modelling (Naimi and Araújo 2016; Naimi et al. 2014; Naimi et al. 2011), environmental epidemiology (Araújo Navas et al. 2016; Hamm et al. 2015b), and for predicting soil properties (Hengl et al. 2015). The ELSA statistic measures the local spatial associations within the same range (between 0 and 1) for both types of data, making the outputs comparable.

We developed a nonparametric bootstrap test based on the ELSA statistic that is useful to test the hypothesis of independence among spatially distributed quantitative (continuous) and qualitative (categorical) data. Our extensive experiments to demonstrate the size and power of the statistic under a range of conditions suggest the usefulness of the statistic for measuring the local spatial association. In addition, we compared the results of our ELSA test based on a Monte Carlo simulation with the well-known Moran’s I test. The ELSA test behaved consistent when the global spatial autocorrelation was either positive or zero. The other LISAs showed inconsistent behaviour, i.e., local Moran’s I and local Geary’s c only showed the expected negative relationship when the
global autocorrelation was zero. The same holds with the p-values generated by the local Moran’s I (Fig. 11 & 12). It has been argued that testing the hypothesis of spatial dependence based on LISAs will give incorrect significance levels in the presence of global spatial association (Anselin 1995). This explains why the tests based on local Moran’s I statistic did not generate the expected p-values when the global spatial autocorrelation was positive. In contrast, our results showed that the ELSA statistic was not sensitive to the presence of the global spatial association that can be considered as an advantage of ELSA over the LISA statistics.

The ELSA calculation for categorical data supports different levels of dissimilarities. This allows evaluation of the dissimilarity between nominal categories in a graded fashion. Several authors in the field of landscape ecology have considered graded differences for measuring landscape patchiness (DeGraaf and Yamasaki 2002; Desrochers et al. 2003). Use of this approach allows transformation from a nominal to an ordered or numerical scale, and provides a foundation for handling attribute distances for categorical data. This is important when quantifying spatial structure for categorical data, since without this aspect all categories are considered equally dissimilar. This approach, however, relies on the subjective evaluation of the class similarity (Ahlqvist and Shortridge 2010), and requires additional data and a guiding theory. Multi-criteria decision making, the analytic hierarchy process, and conjoint analysis are well-known frameworks for evaluating class similarities in categorical maps (Ahlqvist and Shortridge 2010; Schwering 2008). Further studies are needed to characterize and compare these methods, but in this study, we introduced a more general and simple approach to handle class dissimilarities based on a hierarchical scheme of classes. Given additional knowledge about the meaning of class definitions, a more formal evaluation of the categorical dissimilarities can be used to calculate the ELSA statistic for categorical maps.

To measure spatial association, a majority of studies were devoted to the development of statistics for continuous data. In this study, we addressed three of these commonly used statistics and
explored how ELSA relates to them. Although all of these statistics have been used widely as measures of local spatial associations, they quantify different properties. Therefore, the appropriate technique for identification of the spatial association should correspond to the nature of the question concerning dependence/independence (Getis and Ord 1996). Our results confirmed these differences and revealed that ELSA is more related to local Geary’s $c$. Local Geary’s $c$ (like in a variogram) is based on differences between pairs of observations. Analogous to variance, entropy is a measure of dissimilarity and diversity, and their equivalence has been explored and discussed in several studies (e.g. Ebrahimi et al. 1999; Lindley 1956).

Despite the early specific work in geographical analysis (Batty 1974), entropy has been mainly used in other fields, perhaps mostly in the fields of physics and information theory. The most relevant research (to spatial data) in these disciplines introduced several extensions of the entropy measure that apply to calculating the structural complexity or patterns of two-dimensional dynamical systems (Feldman and Crutchfield 2003; Robinson et al. 2011). In recent years, there were several efforts to develop some entropy-based methods for detecting (global) spatial association and patterns of complexity for univariate data (Matilla-García and Marín 2011; Matilla-García et al. 2012; Pham 2010; Ruiz et al. 2010), or as a spatial weighted information measure of global and local spatial association (Karlström and Ceccato 2000), or to discover different forms of local multivariate relationships (Guo 2010). Similar to these methods, ELSA is also based on the entropy measure, but can be considered as a novel approach that offers some unique features, as described in the manuscript.

Together with ELSA, this study introduced the entrogram, an approach for exploring the global spatial structure within the entire area (like a variogram). Such explorations are applied in many fields, such as landscape ecology, geography, and soil science. The entrogram uses the ELSA statistic, and therefore, has the advantage over variogram that can be used for different forms of categorical data (e.g., binary-, multinomial-, or hierarchical-classes) as well as continuous data. The indicator
variogram is a known technique to measure the spatial variability of classes in categorical variables. However, it has been argued that the binary treatment of categorical variables in this technique is an unnecessary oversimplification, and that it should be replaced by ordered measures based on semantic similarity evaluations (Ahlqvist and Shortridge 2010). Semantic variograms (Ahlqvist and Shortridge 2006) were developed based on this concern and provide the capability to consider semantic distance between categories in the calculation. We showed that the entrogram is capable of this as well.

We showed that the entrogram can be used as an exploratory tool to characterise global spatial structure. Variograms, however, can also be used for geostatistical modelling and prediction (kriging). Although this study introduced the entrogram as an exploratory tool, future studies may focus on investigating the use of both ELSA and the entrogram for spatial modelling. Moreover, we believe that the ELSA statistic can be adapted for measuring local temporal autocorrelation, local spatio-temporal autocorrelation and multivariate spatial and spatio-temporal autocorrelation. These might be considered as the areas for future research.

Although raster datasets were used in this study to simplify the illustrations, ELSA can be quantified for other types of spatial data (i.e., spatial points and polygons). Our new R package, elsa, offers the functionalities to quantify ELSA, entrogram (as well as variogram and the other LISA statistics) for all the spatial data types. The software is developed using both the R and C programming languages to speed up the computations. All the functionalities in the package are followed by the help pages where their usage and the relevant details, including some examples, are provided.

8. Conclusion
This paper focused on the development of an entropy-based statistic (ELSA) for the quantification of local spatial association. The ELSA statistic presented in this paper showed to be a robust and reliable method for identifying and testing the degree of spatial association. The method provides the advantage of using one statistic for both continuous and categorical data. This makes comparisons between spatial structures of both types of data possible. Another advantage of ELSA, compared to the LISA statistics, is that it is not sensitive to the presence of the global spatial autocorrelation. In addition, this statistic provides the ability to incorporate both spatial and attribute aspects of spatial association into the statistic for both continuous and categorical data. It can be tested for significance, as demonstrated by a power test in this article. By introducing the 'entrogram' we demonstrated that ELSA can also be used to measure global spatial structure of both continuous and categorical data.

References


Anselin, L. (1993). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. Regional Research Institute, West Virginia University Morgantown, WV


Dacey, M.F. (1968). A review on measures of contiguity for two and k-color maps. In B.J.L. Berry, & D.F. Marble (Eds.), *Spatial analysis: a reader in statistical geography* (pp. 479-495). New Jersey:


James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer


R Development Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria


Tables

Table 1 the level of dissimilarity between pairs of categories in an exemplified land use map with four (sub)categories: 'mixed forest' ($\alpha_1$), 'Coniferous forest' ($\alpha_2$), 'olive groves' ($\alpha_3$), 'vineyards' ($\alpha_4$);

the first two belong to the main category of 'Forests' and the last two are related to 'Agricultural areas'.

<table>
<thead>
<tr>
<th>Categories</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2: Size and power of the ELSA test for continuous, and categorical variables with different degrees of spatial autocorrelation, ranging from no autocorrelation ($\gamma = 0$) to high autocorrelation ($\gamma = 1$); $K$ specifies the number of categories in the categorical variables, and $Ne$ specifies the neighbourhood size for measuring the ELSA statistic.

<table>
<thead>
<tr>
<th>$Ne$</th>
<th>$\gamma = 0$</th>
<th>$\gamma = 0.25$</th>
<th>$\gamma = 0.5$</th>
<th>$\gamma = 0.75$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0.045</td>
<td>0.487</td>
<td>0.882</td>
<td>0.894</td>
<td>0.990</td>
</tr>
<tr>
<td>3</td>
<td>0.048</td>
<td>0.408</td>
<td>0.807</td>
<td>0.916</td>
<td>0.983</td>
</tr>
<tr>
<td>5</td>
<td>0.039</td>
<td>0.398</td>
<td>0.792</td>
<td>0.891</td>
<td>0.977</td>
</tr>
<tr>
<td>10</td>
<td>0.031</td>
<td>0.294</td>
<td>0.734</td>
<td>0.825</td>
<td>0.897</td>
</tr>
<tr>
<td>15</td>
<td>0.012</td>
<td>0.207</td>
<td>0.572</td>
<td>0.650</td>
<td>0.789</td>
</tr>
<tr>
<td>1.5</td>
<td>0.028</td>
<td>0.406</td>
<td>0.787</td>
<td>0.801</td>
<td>0.941</td>
</tr>
<tr>
<td>3</td>
<td>0.017</td>
<td>0.423</td>
<td>0.697</td>
<td>0.787</td>
<td>0.866</td>
</tr>
<tr>
<td>5</td>
<td>0.012</td>
<td>0.412</td>
<td>0.701</td>
<td>0.756</td>
<td>0.840</td>
</tr>
<tr>
<td>10</td>
<td>0.010</td>
<td>0.391</td>
<td>0.639</td>
<td>0.707</td>
<td>0.718</td>
</tr>
<tr>
<td>15</td>
<td>0.002</td>
<td>0.322</td>
<td>0.596</td>
<td>0.650</td>
<td>0.659</td>
</tr>
<tr>
<td>1.5</td>
<td>0.046</td>
<td>0.565</td>
<td>0.705</td>
<td>0.846</td>
<td>0.944</td>
</tr>
<tr>
<td>3</td>
<td>0.005</td>
<td>0.511</td>
<td>0.717</td>
<td>0.827</td>
<td>0.903</td>
</tr>
<tr>
<td>5</td>
<td>0.003</td>
<td>0.495</td>
<td>0.713</td>
<td>0.812</td>
<td>0.892</td>
</tr>
<tr>
<td>10</td>
<td>0.002</td>
<td>0.425</td>
<td>0.601</td>
<td>0.729</td>
<td>0.856</td>
</tr>
<tr>
<td>15</td>
<td>0.000</td>
<td>0.330</td>
<td>0.593</td>
<td>0.698</td>
<td>0.765</td>
</tr>
<tr>
<td>1.5</td>
<td>0.013</td>
<td>0.515</td>
<td>0.741</td>
<td>0.885</td>
<td>0.927</td>
</tr>
<tr>
<td>3</td>
<td>0.011</td>
<td>0.498</td>
<td>0.705</td>
<td>0.822</td>
<td>0.955</td>
</tr>
<tr>
<td>5</td>
<td>0.009</td>
<td>0.413</td>
<td>0.702</td>
<td>0.804</td>
<td>0.926</td>
</tr>
<tr>
<td>10</td>
<td>0.000</td>
<td>0.354</td>
<td>0.619</td>
<td>0.781</td>
<td>0.876</td>
</tr>
<tr>
<td>15</td>
<td>0.000</td>
<td>0.283</td>
<td>0.532</td>
<td>0.773</td>
<td>0.792</td>
</tr>
</tbody>
</table>
### Table A1: The CORINE land cover class definitions

<table>
<thead>
<tr>
<th>CORINE Code</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>Artificial surfaces</td>
<td>Urban fabric</td>
<td>Continuous urban fabric</td>
</tr>
<tr>
<td>112</td>
<td>Artificial surfaces</td>
<td>Urban fabric</td>
<td>Discontinuous urban fabric</td>
</tr>
<tr>
<td>113</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport</td>
<td>Industrial or commercial</td>
</tr>
<tr>
<td>114</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport</td>
<td>Road and rail networks</td>
</tr>
<tr>
<td>115</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport</td>
<td>Airports</td>
</tr>
<tr>
<td>116</td>
<td>Artificial surfaces</td>
<td>Mine, dump and construction sites</td>
<td>Mineral extraction sites</td>
</tr>
<tr>
<td>117</td>
<td>Artificial surfaces</td>
<td>Mine, dump and construction sites</td>
<td>Dump sites</td>
</tr>
<tr>
<td>118</td>
<td>Artificial surfaces</td>
<td>Mine, dump and construction sites</td>
<td>Construction sites</td>
</tr>
<tr>
<td>119</td>
<td>Artificial surfaces</td>
<td>Artificial, non-agricultural vegetated</td>
<td>Green urban areas</td>
</tr>
<tr>
<td>120</td>
<td>Artificial surfaces</td>
<td>Artificial, non-agricultural vegetated</td>
<td>Sport and leisure facilities</td>
</tr>
<tr>
<td>211</td>
<td>Agricultural areas</td>
<td>Arable land</td>
<td>Non-irrigated arable land</td>
</tr>
<tr>
<td>212</td>
<td>Agricultural areas</td>
<td>Arable land</td>
<td>Permanently irrigated land</td>
</tr>
<tr>
<td>213</td>
<td>Agricultural areas</td>
<td>Permanent crops</td>
<td>Vineyards</td>
</tr>
<tr>
<td>214</td>
<td>Agricultural areas</td>
<td>Permanent crops</td>
<td>Fruit trees and berry plantations</td>
</tr>
<tr>
<td>215</td>
<td>Agricultural areas</td>
<td>Permanent crops</td>
<td>Olive groves</td>
</tr>
<tr>
<td>216</td>
<td>Agricultural areas</td>
<td>Pastures</td>
<td>Pastures</td>
</tr>
<tr>
<td>217</td>
<td>Agricultural areas</td>
<td>Heterogeneous agricultural areas</td>
<td>Annual crops/permanent crops</td>
</tr>
<tr>
<td>218</td>
<td>Agricultural areas</td>
<td>Heterogeneous agricultural areas</td>
<td>Complex cultivation patterns</td>
</tr>
<tr>
<td>219</td>
<td>Agricultural areas</td>
<td>Heterogeneous agricultural areas</td>
<td>Land occupied by agriculture</td>
</tr>
<tr>
<td>220</td>
<td>Agricultural areas</td>
<td>Heterogeneous agricultural areas</td>
<td>Agro-forestry areas</td>
</tr>
<tr>
<td>231</td>
<td>Forest and semi natural</td>
<td>Forests</td>
<td>Broad-leaved forest</td>
</tr>
<tr>
<td>232</td>
<td>Forest and semi natural</td>
<td>Forests</td>
<td>Coniferous forest</td>
</tr>
<tr>
<td>233</td>
<td>Forest and semi natural</td>
<td>Forests</td>
<td>Mixed forest</td>
</tr>
<tr>
<td>234</td>
<td>Forest and semi natural</td>
<td>Scrub and/or herbaceous vegetation</td>
<td>Natural grasslands</td>
</tr>
<tr>
<td>235</td>
<td>Forest and semi natural</td>
<td>Scrub and/or herbaceous vegetation</td>
<td>Sclerophyllous vegetation</td>
</tr>
<tr>
<td>236</td>
<td>Forest and semi natural</td>
<td>Scrub and/or herbaceous vegetation</td>
<td>Transitional woodland-shrub</td>
</tr>
<tr>
<td>237</td>
<td>Forest and semi natural</td>
<td>Open spaces with little/no vegetation</td>
<td>Beaches, dunes, sands</td>
</tr>
<tr>
<td>238</td>
<td>Forest and semi natural</td>
<td>Open spaces with little/no vegetation</td>
<td>Bare rocks</td>
</tr>
<tr>
<td>239</td>
<td>Forest and semi natural</td>
<td>Open spaces with little/no vegetation</td>
<td>Sparserly vegetated areas</td>
</tr>
<tr>
<td>240</td>
<td>Forest and semi natural</td>
<td>Open spaces with little/no vegetation</td>
<td>Burnt areas</td>
</tr>
<tr>
<td>241</td>
<td>Wetlands</td>
<td>Inland wetlands</td>
<td>Inland marshes</td>
</tr>
<tr>
<td>242</td>
<td>Water bodies</td>
<td>Inland waters</td>
<td>Water courses</td>
</tr>
<tr>
<td>243</td>
<td>Water bodies</td>
<td>Inland waters</td>
<td>Water bodies</td>
</tr>
</tbody>
</table>

---
Fig. A1. Hierarchical scheme of different classes in the CORINE land cover map (Fig. 7 in the main text)
**Fig. A2.** The level of dissimilarity between pairs of classes in the CORINE land cover map
**Figure Captions**

**Figure 1.** The behaviour of ELSA statistic (E) and its two terms at site i (the centre of the window):

- $E_{ai}$ (first term in ELSA refers to attribute distance or dissimilarity between the site i and the its neighbour locations),
- $E_{ci}$ (second term in ELSA, refers to composition of categories within the local area);
- (a) is a map with one category, (b, and c) are binary maps, and (d, and e) are maps with three categories.

**Figure 2.** A hierarchical way of presenting the classes in an exemplified categorical map with 3 levels of categories; the numbers in the boxes indicate d (dissimilarity) of the relevant pairs of classes.

**Figure 3.** A flow diagram showing the procedure of finding the optimum number of categories for categorizing continuous spatial data; (a) an iterative categorization procedure taking different number of categories, starts from 2 categories and continues by considering one more category at each iteration until a convergence threshold (e.g., 0.005) is reached; (b) calculating the $\rho$ correlation coefficient between the continuous values and each categorical variable; the convergence is defined as the difference between the $\rho$ coefficients of the current and previous iterations; (c) taking the standard error of the $\rho$ coefficients is calculated; (d) the optimum number of categories would be the lowest number for which the $\rho$ coefficient is within one standard error of the highest $\rho$ coefficient; (e) the graph representing the $\rho$ coefficients over different iterations and the optimum number of categories (dashed red line); the grey colour represents the standard error of the $\rho$ coefficients.

**Figure 4.** Four variogram models with different nugget effects ranging between 0 and 10, used to simulate the spatial surfaces with different degrees of spatial autocorrelation.

**Figure 5.** One realization (out of 999) of the simulated continuous and categorical maps with various levels of spatial autocorrelation to calculate the power and size of the ELSA statistic; $\gamma$ is the ratio.
between the partial sill to the sill in the variogram model used to simulate the maps (Figure 4) that controls the level of spatial autocorrelation ranging from no autocorrelation ($\gamma=0$), and high autocorrelation ($\gamma=1$); $K$ specifies the number of classes in the categorical maps.

Figure 6. ELSA for a land cover map in the south of Spain; (a) the land cover map including six classes with the same level of dissimilarity between different categories; three cells within their five km neighbourhoods are specified as A, B, C; (b) the ELSA statistic for the land cover map; the values of ELSA for the three cells are represented on top

Figure 7. Synthetic land cover map including four classes which distributions are controlled into four equal zones (a); levels of dissimilarity between pairs of classes in a hierarchical view (b)

Figure 8. The ELSA map (a) and the Mean ELSA statistic in four zones of the region (b); the region is divided into four zones including Z1 (upper-left), Z2 (upper-right), Z3 (lower-right) and Z4 (lower-left); the boxplot (c) represents the distribution of ELSA values at grid cells over different zones

Figure 9. CORINE Land cover map from the central Spain; a three-digit code is used to define each land cover class (a); three randomly selected points around which the circle specifies a 5 km of their neighbors (b)

Figure 10. ELSA map calculated based on the CORINE land cover map in Fig. 7 (a); the ELSA value at the three specified locations (b)

Figure 11. Comparing local indicators of spatial association for a continuous raster map with positive global spatial autocorrelation; the panel contains 12 map/graph/table including the simulated raster map, a Moran scatterplot showing the degree of global spatial autocorrelation, four maps of local Moran’s I, local Geary’s c, local $G^+$, and ELSA, 2 maps representing the p-values that indicate the
level of significance for local spatial association based on local Moran’s I, and ELSA, respectively, and
3 graphs representing the relationship between ELSA on y axis and local Geary’s c, local Moran’s I,
and $G_1^*$ statistics on x axis, and finally a table representing the pairwise spearman correlation
coefficients between different local indicators of spatial association; the symbol ‘***’ indicates that
the correlation coefficient is significant at the level of 99.9%

**Figure 12.** Comparing local indicators of spatial association for a continuous raster maps with no
global spatial autocorrelation; the different components of the figure are declared in Fig. 11.

**Figure 13.** Comparing variogram and entrogram for three simulated binary categorical maps with
different levels of spatial autocorrelation ($\phi = 0, 3, 8$ from left to right); the second and third rows
display the corresponding theoretical and empirical variograms, respectively, and the last row
displays the entrograms

**Figure 14.** Two categorical maps including four classes (first row) and their corresponding
entrograms (second row)

**Figure 15.** Comparing variogram and entrogram; the first row displays the 5 simulated continuous
fields with different levels of spatial autocorrelation ($\phi = 0, 10, 30, 10, 10$ from left to right), and
varying parameterisation in the variogram model, nugget ($C_0$) and partial sill ($C$) (as specified in the
figure); the second row displays the corresponding variograms and the third row displays the
entrograms
**Figure 1**
Figure 2.
Figure 3.
Figure 4.
<table>
<thead>
<tr>
<th></th>
<th>Continuous</th>
<th>Categorical (K=2)</th>
<th>Categorical (K=3)</th>
<th>Categorical (K=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = 0$</td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
</tr>
<tr>
<td>$\gamma = 0.25$</td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
</tr>
<tr>
<td>$\gamma = 0.5$</td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
</tr>
<tr>
<td>$\gamma = 0.75$</td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
</tr>
<tr>
<td>$\gamma = 1$</td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
<td><img src="#" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 5.
Figure 6
Figure 7.
Figure 8.
Figure 9.
Figure 10.
Figure 12.
Figure 13.
Figure 14.
<table>
<thead>
<tr>
<th>Continuous-Kernels</th>
<th>No Spatial Autocorrelation</th>
<th>Model: Spherical $\beta = 10, \lambda_0 = 0, \beta = 10$</th>
<th>Model: Spherical $\beta = 30, \lambda_0 = 0, \beta = 10$</th>
<th>Model: Gaussian $\beta = 10, \lambda_0 = 3, \beta = 7$</th>
<th>Model: Gaussian $\beta = 10, \lambda_0 = 5, \beta = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1" alt="Continuous-Kernels" /></td>
<td><img src="image2" alt="Continuous-Kernels" /></td>
<td><img src="image3" alt="Continuous-Kernels" /></td>
<td><img src="image4" alt="Continuous-Kernels" /></td>
<td><img src="image5" alt="Continuous-Kernels" /></td>
</tr>
</tbody>
</table>

Figure 15.